Wavelet-based 2D Multifractal Spectrum with Applications in Analysis of Digital Mammography Images

Pepa Ramírez and Brani Vidakovic

July 26, 2007

Abstract

Breast cancer is the second leading cause of death in women in the United States and at present, mammography is the only proven method that can detect minimal breast cancer. On the other hand, many medical images demonstrate a certain degree of self-similarity over a range of scales. The Multifractal spectrum (MFS) summarizes possibly variable degrees of scaling in one dimensional signals and has been widely used in fractal analysis. In this work, we develop a generalization to two dimensions of MFS and use dynamics of the scaling as discriminatory descriptors to do classification of mammographic images to benign and malignant. Methodology we propose was tested using images from the University of South Florida Digital Database for Screening Mammography (DDSM) (Heat et al. [8]).

1 Introduction

In the United States, breast cancer is the second leading cause of death in women. One out of eight women will develop breast cancer in their lifetime. Studies have indicated that early detection and treatment improve the chances of survival for breast cancer patients (Curpen et al. [5], Smart et al [25]). At present, mammography is the only proven method that can detect minimal breast cancers. However, the radiological interpretation of mammograms is a difficult task since the mammographic appearance of normal tissue is highly variable and complex structures in appearance and signs of early disease are often small or subtle. That is the main cause of many misdiagnoses that can be attributed mainly to human factors (Martin et al [20], Kalisher

[12]) and this explains that $10-30\%$ of cancers which could have been detected are missed. On the other hand, a high percentage of patients called back at screening turn out not to have cancer. Because of these concerns much research has been devoted to developing methods for classifying suspicious areas of mammography tissue. This could help radiologists to improve the efficacy of screening programs and avoid unnecessary biopsies.

Wavelet techniques have been proven to be indispensable for image processing, in particular when dealing with medical images such as mammograms. Mallat's multiresolution analysis (see Vidakovic [29]) decomposes an image into a set of approximation coefficients (low frequency components) and the scale dependent hierarchy of detail coefficients (high frequency components). Standard tensor product orthogonal wavelet transformation of an image results in three sets of generated detail coefficients: diagonal, horizontal and vertical. Numerous references can be found in the literature in which wavelets are applied to mammogram images. For example, in Yoshida et al. [31], a wavelet transform technique was applied to detect clustered microcalcifications. In Zheng et al. [32], a wavelet-based image-enhancement method is employed to enhance microcalcification clusters for improved detection. The authors there use a neural network methodology for detecting microcalcification clusters. In Bruce and Adhami [3], the wavelet modulus-maxima method is utilized for the extraction of mammographic lesion shape features. Moreover, a linear discriminant analysis is used to discriminate the features. This wavelet transform modulus maxima method (or WTMM) is generalized to multifractal image analysis in Arnéodo et al. [1]. Various wavelet basis were used in Lado et al. [16] to detect clustered microcalcifications in digital mammograms.

Fractality is a concept pervasive in the medical field; many medical images demonstrate a certain degree of self-similarity over a range of scales, lending to development of algorithms based on fractal analysis of those images (see Chen et al. [4] and Kuklinski [15]). For example, fractality was used to detect breast cancer, in for instance Priebe et al. [23], Kestener et al. [30], and Bocchi et al. [2]. In Chen et al. [4], a pattern recognition technique based on features derived from the factal description of mammograms is developed. In Kuklinski [15], the authors use a wavelet transform modulus maxima method generalized to the 2-dimensional case. They combine this approach with a multifractal analysis which enables them to detect tumors as well as microcalcifications. And in Bocchi et al. [2] the fractional Brownian motion model is combined with neural classifiers for detection and classification of microcalcifications.

The Multifractal spectrum (MFS) summarizes possibly variable degrees of scaling in signals; in the case of fractals, scaling refers to the propagation of energy when the signals or images are inspected as various resolutions. The dynamics of the scaling can be used as discriminatory descriptors, thus, multifractality provides an additional window through which to look at the data and makes inference not possible with standard statistical approaches. There are many examples of use of multifractals in the literature. In Kestener et al. [30], long range correlations and wavelet-based multifractality were used for tissue classification in digitized mammograms to support clinical diagnosis. In Moloney et al. [19], the MFS is used to analyze the pupillary behavior of older adults and discriminate between patients with various ocular acuity.

In this work, we generalize the concept of Multifractal spectrum as it was defined in Gonçalves et al. [22] to the 2-dimensional case and use some of its descriptors in classification of mammographic images to benign and malignant. Several mammogram classification methods can be found in literature but with other descriptors. The examples are linear, neural and support vector machines (see Mavroforakis et at. [18]) and independent component analysis classifiers (Koutras et al. [14]). In particular, we tested many of them Section 4 applied on descriptors of 2D-MFS. Our data set coming from real life data base is composed of digitized mammograms to which we apply wavelet transformation and compute their respective 2D MFS. Instead of focusing on microcalcifications as previous researches did, we explore the background of the image. From the obtained results we conclude that there exist significant differences in the background of benign and malignant mammograms in terms of multifractality.

The paper is organized as follows. In Section 2 a theoretical background of wavelets and the definition and key properties of fractional Brownian motion in one and two dimensions are briefly reviewed. In addition, the 2 dimensional Multifractal spectrum is defined and some of its properties are illustrated on the example of 2D fractional Brownian motion MFS. Section 4 deals with a real-life application of the proposed methodology: classification of digitized mammograms. In Section 5 we provide conclusions and delineate some possible directions for future research.

2 Multifractal Analysis

The fractional Brownian motion is a non-stationary process whose sample paths exhibit a homogeneous degree of smoothness or regularity. And for many applications, this regularity may be too homoge-

neous. In particular, one may want the sample paths of a process that exhibits differing degrees of regularity as a function of time. Processes with fractal characteristics that exhibit such degree of complexity are often referred to as *multifractals*. Multifractal analysis is concerned with describing the local singular behavior of functions in a geometrical and statistical fashion. It was first introduced in the context of turbulence and applied in many other contexts such as DLA patterns research, earth quake distribution analysis, signal processing and internet data traffic modelling. For an introduction to multifractals, see Riedi, [26].

Multifractal processes exhibit patterns of locally varying scaling behavior similar to that encountered in some real data sets. They usually exhibit a prevalent scaling behavior, but a multitude of other scalings may also be present although occurring much less frequently. Since multifractal processes are in general non-stationary, standard approaches in time series analysis such as Fourier transform are not appropriate because the Fourier transform is not localized in time. In order to study the varying local properties of multifractal processes, tools able to localize information in time and frequency are appropriate. Given that wavelets are local in both frequency/scale (via dilations) and in time (via translations), the wavelet defined multiscale analysis is convenient in this setting. For a detailed study of multifractal processes we refer the reader to Riedi [24] and Morales [10].

2.1 Background on wavelets

The discrete wavelet transform is representing a 1-D real signal $X(t)$ in terms of shifted and dilated versions of a wavelet (or *mother*) function $\psi(t)$ and shifted versions of a low-pass scaling (or *father*) function $\phi(t)$. For specific choices of the wavelet and scaling functions, an orthonormal basis can be formed from the atoms

$$\begin{aligned} \psi_{j,k}(t) &= 2^{j/2}\psi(2^j t - k), \\ \phi_{j,k}(t) &= 2^{j/2}\phi(2^j t - k), \quad j,k \in \mathbb{Z} \end{aligned}$$

For instance, the signal $X(t)$ can be represented by wavelet basis as

$$X(t) = \sum_k c_{J_0,k}\phi_{J_0,k}(t) + \sum_{j=J_0}^{\infty}\sum_k d_{j,k}\psi_{j,k}(t),$$

with

$$\begin{aligned} d_{j,k} &= \int X(t)\psi_{j,k}(t)dt, \\ c_{j,k} &= \int X(t)\phi_{j,k}(t)dt \end{aligned} \tag{2.1}$$

Here, J_0 indicates the coarsest scale or lowest resolution of analysis, and larger j correspond to higher resolutions of the analysis (for a detailed wavelets theory, see Vidakovic [29]). In practice, many signals are multidimensional. Examples include measurements in geophysics, medicine, astronomy, economics, and so on. The wavelet transform are readily generalized to multidimensional case. Since we are interested in the wavelet transforms of medical images, the generalization we show is for the 2-dimensional case. The 2-D wavelet basis functions are constructed via translations and dilations of a tensor product of univariate wavelets and scaling functions:

$$\begin{aligned} \phi(x_1, x_2) &= \phi(x_1)\phi(x_2) \\ \psi^h(x_1, x_2) &= \phi(x_1)\psi(x_2) \\ \psi^v(x_1, x_2) &= \psi(x_1)\phi(x_2) \\ \psi^d(x_1, x_2) &= \psi(x_1)\psi(x_2) \end{aligned} \tag{2.2}$$

The symbols h, v, d in (2.2) stand for horizontal, vertical and diagonal directions, respectively. For technical reasons we consider L^1-normalization of wavelets atoms instead of standard L^2 normalization, which expression for $\psi_{j,\mathbf{k}}, \phi_{j,\mathbf{k}}$ is,

$$\begin{aligned} \phi_{j,\mathbf{k}}(\mathbf{x}) &= 2^{2j}\phi(2^j x_1 - k_1, 2^j x_2 - k_2) \\ \psi^i_{j,\mathbf{k}}(\mathbf{x}) &= 2^{2j}\psi^i(2^j x_1 - k_1, 2^j x_2 - k_2) \end{aligned}$$

for $i = h, v, d$ and where $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$, and $\mathbf{k} = (k_1, k_2) \in \mathbb{Z}^2$.

Then, any function $f \in \mathbb{L}_2(\mathbb{R}^2)$ can be represented as

$$f(\mathbf{x}) = \sum_{\mathbf{k}} c_{j_0\mathbf{k}}\phi_{j_0,\mathbf{k}}(\mathbf{x}) + \sum_{j \geqslant j_0}\sum_{\mathbf{k}}\sum_{i} d^i_{j,\mathbf{k}}\psi^i_{j,\mathbf{k}}(\mathbf{x}) \tag{2.3}$$

where the wavelet coefficients are given by

$$d^i_{j,\mathbf{k}} = 2^{2j}\int f(\mathbf{x})\psi^{\mathbf{i}}(2^{\mathbf{j}}\mathbf{x} - \mathbf{k})\mathrm{d}\mathbf{x} \tag{2.4}$$

Expression (2.3) for $f(\mathbf{x})$ can be generalized to d dimensions so that if $f \in \mathbb{L}_2(\mathbb{R}^d)$, then

$$f(\mathbf{x}) = \sum_{\mathbf{k}} c_{j_0\mathbf{k}}\phi_{j_0,\mathbf{k}}(\mathbf{x}) + \sum_{j \geqslant j_0}\sum_{\mathbf{k}}\sum_{i=1}^{2^d-1} d^i_{j,\mathbf{k}}\psi^i_{j,\mathbf{k}}(\mathbf{x})$$

where $\mathbf{k} = (k_1, \ldots, k_d) \in \mathbb{Z}^d, \mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$, and in L^1-norm,

$$\begin{aligned} \phi_{j,\mathbf{k}}(\mathbf{x}) &= 2^{jd} \prod_{i=1}^{d} \phi(2^j x_i - k_i) \\ \psi^l_{j,\mathbf{k}}(\mathbf{x}) &= 2^{jd} \prod_{i=1}^{d} \xi(2^j x_i - k_i) \end{aligned}$$

with $\xi = \phi$ or ψ, but not all $\xi = \phi$. The index l corresponds to one of the $2^d - 1$ possible directions. The 2-dimensional multifractal wavelet spectra will be defined using the wavelet coefficients $d^i_{j,\mathbf{k}}$, along the scale index j. We assume that the mother wavelet ψ has $\mathcal{R}$ vanishing moments, that is, $\int x^r \psi(x) dx = 0, r = 0, \ldots, \mathcal{R}$, because the decorrelation property of wavelet coefficients depends upon this assumption.

2.2 2D Multifractal spectrum

In Gonçalves [22], it is shown how the oscillatory or scaling behavior of a process carries over into the local scaling properties of its wavelets coefficients $d_{j,k}$ (2.1), under assumption that the wavelet is more regular than the process. The following *local singularity strenght measure* in 2 dimensions can be defined using wavelets,

$$\alpha^i(\mathbf{t}) := \lim_{\mathbf{k}2^{-\mathrm{j}} \to \mathbf{t}} -\frac{1}{j} \log_2 |d^i_{j,\mathbf{k}}| \tag{2.5}$$

where $\mathbf{k}2^{-j} \to \mathbf{t}$ means that $\mathbf{t} = (t_1, t_2) \in [2^{-j}k_1, 2^{-j}(k_1 + 1)] \times [2^{-j}k_2, 2^{-j}(k_2+1)]$ for $\mathbf{k} = (k_1, k_2)$ and $j \to \infty$. Smaller $\alpha(\mathbf{t})$correspond to larger oscillations in X and thus to more singularity at time $\mathbf{t}$. The index i in (2.5) corresponds to one of three directions in detail spaces of 2D wavelet transform, horizontal (h), vertical (v) or diagonal (d). Typically, a process will possess many different singularity strengths. The frequency (in $\mathbf{t}$) of occurrence of a given singularity strength α is measured by the *2D multifractal spectrum*, defined for each direction $i = d, h, v$ as

$$\begin{aligned} f^i(\alpha) &:= \lim_{\epsilon \to 0} \lim_{j \to \infty} \frac{1}{j} \log_2 M^i_j \\ M^i_j &:= 2^j \#\{\mathbf{k} : 2^{\mathrm{j}(\alpha+\epsilon)} \leqslant |\mathrm{d}^{\mathrm{i}}_{\mathrm{j},\mathbf{k}}| \leqslant 2^{\mathrm{j}(\alpha+\epsilon)}\}. \end{aligned} \tag{2.6}$$

For $\mathbf{k} \in \{0, \ldots, 2^{\mathrm{j}} - 1\} \times \{0, \ldots, 2^{\mathrm{j}} - 1\}$, f^i takes values between -1 and 0. Smaller $f^i(\alpha)$ means that "fewer" points $\mathbf{t}$ behave with strength $\alpha(\mathbf{t}) \simeq \alpha$.

2.3 2D Multifractal formalism

The 2D multifractal spectrum f^i defined in (2.6) is very hard to calculate. A simpler approach makes use of the theory of large deviations (see Ellis, [7]), where f^i would be interpreted as the rate function of a Large Deviation Principle: f^i measures how frequently (in $\mathbf{k}$) the observed $(-1/j)\log_2|d^i_{j,\mathbf{k}}|$ deviate from the "expected value" α_0 in scale j. In our 2D context, it corresponds to studying the scaling behavior of the moments of the wavelets coefficients (2.4). For every direction i, the *partition function* is defined,

$$T^i(q) := \lim_{j\to\infty} \log_2 \mathbb{E}|d^i_{j,\mathbf{k}}|^q \tag{2.7}$$

It describes limiting behaviour of qth moment of a typical wavelet coefficient d^i_j, k from the level j and direction i. The *multifractal formalism* posits that the multifractal spectrum can be calculated by taking the Legendre transform of the corresponding log moment generating function (Riedi et al. [27])

$$f^i(\alpha) = f^i_L(\alpha) := \inf_q [q\alpha - T^i(q)]. \tag{2.8}$$

It can be shown that $f^i_L(\alpha) = q\alpha - T^i(q)$ at $\alpha^i = T'^i(q)$ provided $T''^i(q) < 0$.

2.4 Fractional Brownian motion in two dimensions

In this section we review the definition of the fractional Brownian motion (fBm), a very popular model in signal and image processing for description of data that scale in a regular fashion. The fBm has proved useful for modeling various physical phenomena involving long-range dependence, and regular self-similarity. It is a natural extension of the well-known Brownian motion (see Mandelbrot and Ness [21]). It is a Gaussian, zero-mean nonstationary process $B_H(t)$, where $H \in (0,1)$ is known as the Hurst parameter. If $H = 1/2$ the standard Brownian motion is recovered. The autocovariance function fully describes the process. As it can be seen, the process is nonstationary since the autocovariance function is not a function of $|t-s|$

$$E\left(B_H(t)B_H(s)\right) = \frac{\sigma^2}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}) \tag{2.9}$$

The variance of the process is given by

$$var\left(B_H(t)\right) = \sigma^2|t|^{2H}, \tag{2.10}$$

where $\sigma^2 = \Gamma(1-2H)(cos\pi H)/(\pi H)$. As the process is nonstationary, it does not possess a spectrum in the usual sense. However, since it has stationary increments the average spectrum is defined as

$$\mathcal{S}_{B_H}(\omega) = \frac{\sigma^2}{|\omega|^{2H+1}}. \tag{2.11}$$

Because of this power law behavior, the fBm is an appropriate model for various $1/f$-type processes. The self-similary property is captured by

$$B_H(at) \stackrel{d}{=} a^H B_H(t) \tag{2.12}$$

where $\stackrel{d}{=}$ means equality in distribution.

The definition of the fractional Brownian motion can be extended to higher dimensions along the lines of Lévy [17], where the generalization of Brownian motion to multiple dimensions was first considered. A 2-D fBm $B_H(\mathbf{t})$, for $\mathbf{t} \in [0,1] \times [0,1]$ and $H \in (0,1)$, is a process with stationary zero-mean Gaussian increments and where the autocovariance function is given by,

$$E\left(B_H(\mathbf{t})\mathrm{B}_H(\mathbf{s})\right) = \frac{\sigma^2}{2}(\|\mathbf{t}\|^{2H} + \|\mathbf{s}\|^{2H} - \|\mathbf{t}-\mathbf{s}\|^{2H}) \tag{2.13}$$

where $\|\cdot\|$ is the usual Euclidean norm in $\mathbb{R}^2$. The variance of such a process is

$$var(B_H(\mathbf{x})) = \sigma^{\mathbf{2}}\|\mathbf{x}\|^{2H},$$

and the average power spectrum of an 2-D fBm is

$$S_{B_H}(\boldsymbol{\omega}) \propto \|\boldsymbol{\omega}\|^{-2H-2},$$

or equivalently,

$$S_{B_H}(\omega_1, \omega_2) \propto \frac{1}{(\omega_1^2 + \omega_2^2)^{H+1}}.$$

The index H corresponds to the Hurst exponent; higher exponents H correspond to a more regular fBm surfaces (as Figure 1 shows).

The wavelet coefficients of an 2D fBm are

$$d^i_{j,\mathbf{k}} = 2^{2j} \int B_H(\mathbf{x})\psi^{\mathbf{i}}(2^{\mathbf{j}}\mathbf{x} - \mathbf{k})\mathrm{d}\mathbf{x} \tag{2.14}$$

where the integral is taken over $\mathbb{R}^2$ and $i = d, v$ or h. They sport the following properties (proofs are given in the Appendix I):

P1. Isotropy: $d^i_{j,\mathbf{k}} \stackrel{d}{=} d^i_{j,0}, \forall \mathbf{k}$

P2. Gaussianity: $d^i_{j,\mathbf{k}} \sim N(0, \sigma_{\psi^i} 2^{2jH})$, where σ_{ψ^i} is a constant depending only on the wavelet function ψ^i.

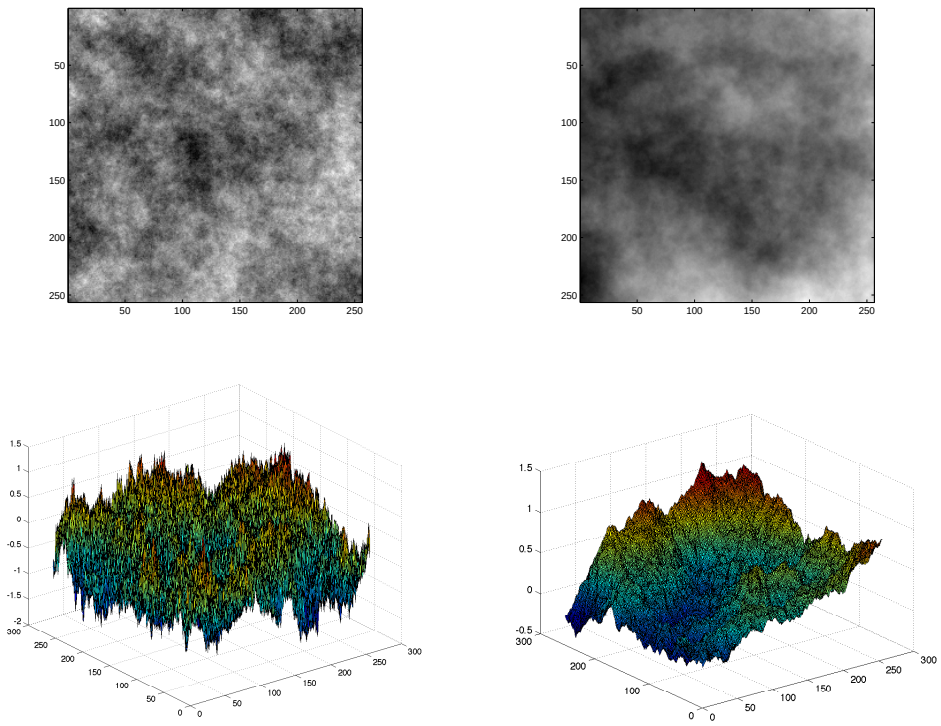

Figure 1: FBm surfaces (256×256) with $H = 0.3$ and $H = 0.7$ respectively. In the top panels, $B_H(\mathbf{x})$ is coded using 32 grey levels from white (min B_H) to black (max B_H).

P3. Almost decorrelation: $E(d^i_{j,\mathbf{k}} d^i_{j',\mathbf{k}'}) \sim \|\mathbf{k}-\mathbf{k}'\|^{2(H-\mathcal{R})}$, where $\mathcal{R}$ is the number of vanishing moments of the mother wavelet ψ, and

P4. Scaling: $d^i_{j,\mathbf{k}} \stackrel{d}{=} 4^{jH} d^i_{1,\mathbf{k}}$

Because of $P2$ and $P3$, it will be assumed that the fBm wavelet coefficient are exactly uncorrelated and hence independent. It is not an unreasonable assumption (see Flandrin [11]). Furthermore, since the moments of order $q < -1$ of a Gaussian process are infinite, either $P2$ or $P4$ yields

$$T^i(q) = \begin{cases} 2qH & q > -1 \\ -\infty & q \leqslant -1 \end{cases}$$

and thus that

$$f^i(\alpha) = f^i_L(\alpha) = \begin{cases} \infty & \alpha < H \\ 0 & \alpha = H \\ H-\alpha & \alpha > H \end{cases}$$

3 Wavelet-based estimator

We discuss in this section wavelet-based estimation of the 2D multifractal spectrum (2.6). Given a realization of the 2D fBm of size $2^J \times 2^J$, and using the stationarity of the wavelets coefficients $\{d^i_{j,(k_1,k_2)}, i = d,h,v; j = J_0,\ldots,J-1, k_1,k_2 = 0,\ldots,2^j-1\}$, the sample counterpart of $\mathbb{E}|d^i_{j,\mathbf{k}}|^q$ is

$$\widehat{S}^i_j(q) := \frac{1}{2^j} \sum_{k_1=0}^{2^j-1} \left(\frac{1}{2^j} \sum_{k_2=0}^{2^j-1} |d^i_{j,(k_1,k_2)}|^q \right) \tag{3.1}$$

for $q > -1$. The partition function can then be estimated as the power-law exponent of the variation of $\hat{S}^i_j(q)$ versus scale 2^{-j}. By linear regression of $\log_2 \hat{S}^i_j(q)$ on j between scales j_1 and j_2 we get

$$\widehat{T}^i(q) := \sum_{j=j_1}^{j_2} a_j \log_2 \widehat{S}^i_j(q), \tag{3.2}$$

where the regression weights a_j must verify the two conditions $\sum_j a_j = 0$ and $\sum_j j a_j = 1$ (Delbeke and Abry [6]). Thus, we can estimate $f^i(\alpha)$ though a local slope of $\widehat{T}^i(q)$ at values

$$\widehat{\alpha}^i(q_l) = [\widehat{T}^i(q_{l+1}) - \widehat{T}^i(q_l)]/q_0, \quad q_l = l q_0$$

as

$$\widehat{f^i}(\alpha^i(q_l)) = q_l\alpha^i(q_l) - \widehat{T}^i(q_l).$$

Multifractal spectra can be found even for monofractal processes, the spectra generated from monofractal processes are ramp-like with a dominant (modal) irregularity corresponding to the theoretical Hurst exponent (see Riedi [24]). Figure 2 depicts the 2D MFS for the simulated 2D fBm showed in Figure 1. Notice how the maximum of every $f^i(\alpha)$ is attained close to $\alpha = 0.3$ and $\alpha = 0.7$, and deviations from the exact values can be attributed to discretization or small number of dyadic levels.

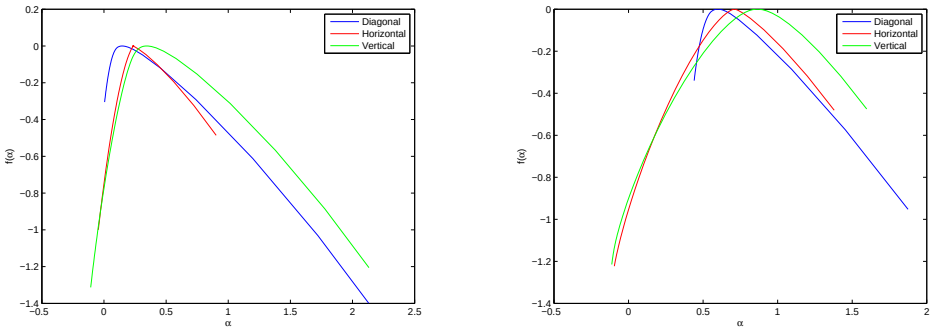

Figure 2: 2D MFS associated to the 2D fBm of Fig.1

4 An application in analysis of Digital Mammography images

In this section we provide an application of the previously defined 2-D Wavelet-based multifractal spectrum to the classification of mammography images. We classify images as benign or malignant, analyzing the background of the image.

4.1 Description of the data

The collection of images we analyzed was obtained from the University of South Florida's Digital Database for Screening Mammography (Heat et al. [8]).
(`http://marathon.csee.usf.edu/Mammography/Database.html`).
The DDSM is described in details in Heat et al. [9]. Images

containing suspicious areas have associated pixel-level "ground truth" information about the locations and types of suspicious regions. We selected a set of cases (studies) from the DDSM from volumes 6 and 7. Each case contains four mammograms (two for each breast, the craniocaudal (CC) and mediolateral oblique (MLO) projections) from a screening exam. We analyzed the data from 5 normal cases and 5 malignant cases, each containing calcifications.

The images were scanned on either a HOWTEK 960 or HOWTEK MultiRAD 850 digitizer with a sample rate of 43.5 microns per second at 12 bits per pixel. They were stored in a format using lossless JPEG compression. However, even with the compression, each image file is quite large because the films were scanned with resolution between 42 and 100 microns. The source code for the program used to compress, as well as the program used to uncompress the images are available to download from the web site.

Each image was divided in non-overlapping subimages, each of size 256×256. Our data set contains 148 subimages, 74 from normal cases and 74 from malignant cases. Notice that some of these subimages coming from a malignant case do not contain the calcification or even tissue close to this.

Figure 4 shows two mammograms, the mammogram on left panel corresponds to a 47 years old woman and it is an example of normal case; the mammogram on right panel comes from a 50 years old patient with a single abnormality. The lesion is a calcification type pleomorphic distribution clustered. The tumor was localized around the coordinates $(800, 2600)$.

4.2 Methodology and results

For every subimage we compute the 2-D multifractal spectrum resulting in three multifractal spectra. Figure 5 shows the 2-D MFS for one of these subimages.

In order to classify our data, we computed, for every MFS in each direction, the measures defined in (Jongphil and Brani paper). These are the Hurst exponent, the left and right tangent when $f(\alpha) = -0.2$, the left and right slope with the Hurst exponent, the bandwidth (or distance between α_1 and α_2 where $f(\alpha_1) = f(\alpha_2) = -0.2$), and the values of α_1 and α_2. For instance, for the previously computed MFS for the vertical direction

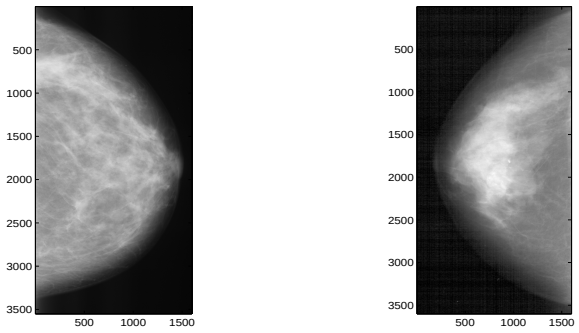

Figure 3: Examples of normal and malignant mammograms

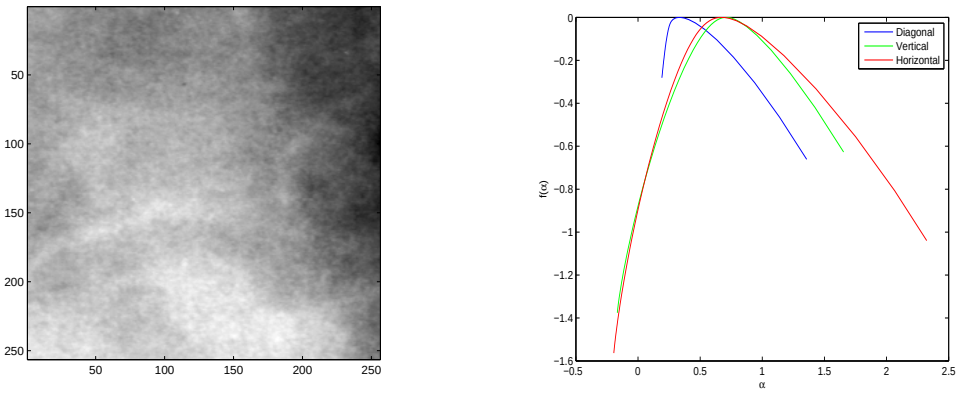

Figure 4: Normal breast tissue and corresponding 2-D MFS.

these quantities were $(0.71, 1.2, -0.7, 0.64, -0.46, 0.73, 0.40, 1.14)$, as Figure 5 shows.

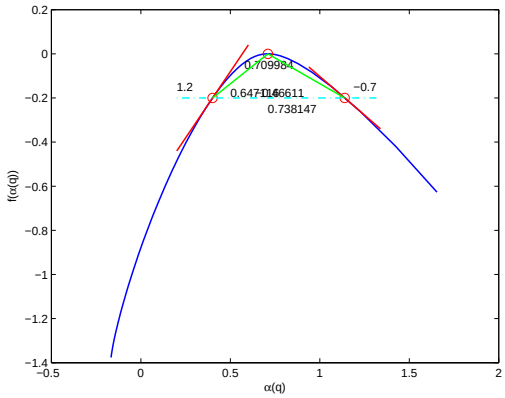

Figure 5: Possible discriminant measures for a given MFS

We computed these eight measures for the MFS associated to every subimage in the sample. Then, for each direction, diagonal, horizontal and vertical we observed the scatterplot of normal versus cancer breasts for all pairs of measures. We did not find any significant difference in the horizontal and vertical directions. However, in the diagonal one, we did. Figure 6 depicts some scatterplots for four different pairs of measures associated to the diagonal MFS. We can see how normal mammographies use to present a smaller Hurst exponent than the malignant ones which means more irregularity in the tissue. We can also observe how the MFS associated to normal cases tend to be slightly shifted to the right, that is, the values α_1, α_2 are smaller for the normal mammographies.This implies larger oscillations, or lower degree of smoothness as compared to cancer cases. We can see, however that the bandwidth seems to be similar in both cases.

In order to classify the patients as with or without a malignant tumor, we tried several discrimination and classification methods. For a detailed explanation of every method, see Hastie et al. [13]. We will briefly describe them and comment the obtained results. The prediction error was estimated by cross-validation. Firstly, we used the 50% of data to fit the model, and the rest to test it. Later, we used the 85% as the training set and the 15% as the sample to test the model. We did it in an iterative way, repeating

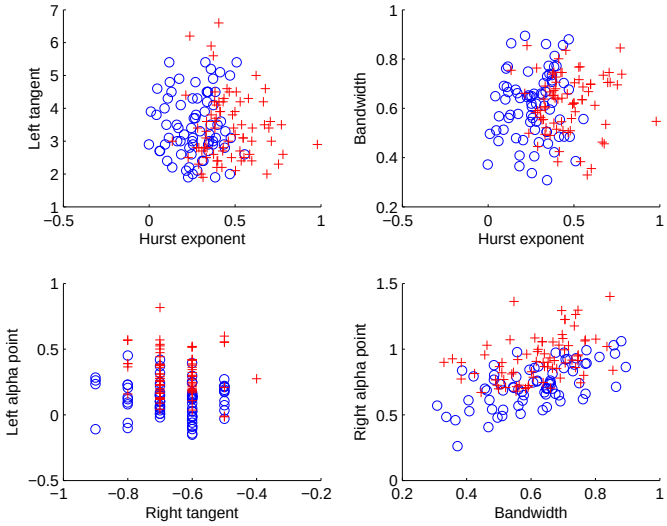

Figure 6: Scatterplots for 4 pairs of measures associated to the diagonal MFS. The symbols denote: *blue circles* for normal mammographies; *red squares* for malignant mammographies.

the experiment 10000 times, so the prediction errors are averaged errors. Tables 1-4 show theses values for different classification methods.

To begin with, we tried classical linear and quadratic discriminant analysis. Both methods are designed for situations where the measurements from each group have a multivariate normal distribution. From tables 1 and 2 we can see that the measures that best discriminate with these methods are the pairs, (left tangent, left α) and (right tangent, left α). Figure 7 depicts the misclassified data when a lineal and quadratic discriminants are applied to the pairs of measures (Hurst exponent, right α) and (Hurst exponent, bandwidth). In this case, the sample, the training and the test sets coincide. The misclassification errors were 0.2568 and 0.2365 respectively.

Sometimes the normality assumption is not valid and in these cases a nonparametric classification procedure may be more appropriate. Next approach we undertook was based on decision trees. They are sets of simple rules like "if the Hurst exponent is less than 0.5, classify the observation as normal". Decision trees do not require any assumptions about the distribution of the mea-

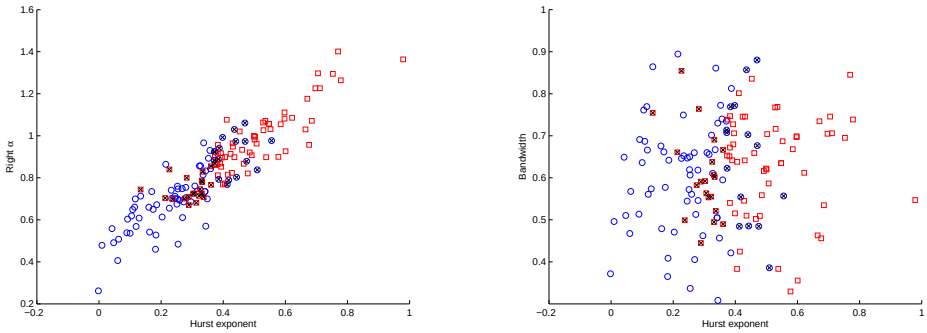

Figure 7: Classification by lineal and quadratic discriminants. The symbols denote: *blue circles* for normal mammographies; *red* crosses for malignant mammographies. The crossed observations are misclassified.

surements in each group. Figure 8 shows how the method separates the plane into regions divided by lines, and assigned different regions to normal and cancer cases. This method performed pretty well when we considered the whole sample as the training and the test as well. In that case, the pair of measures that best discriminated was (Hurst exponent, left tangent), the misclassification error was 0.0946, but when we did cross-validation the results were slightly worse than with the classical methods, as tables 1 and 2 show.

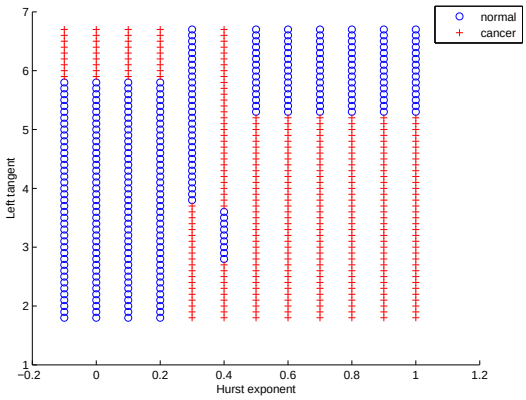

Figure 8: Plane regions assigned to each case, normal and cancer, by the decision tree method.

Next method we tried was the support vector machines or SVM, (see Hastie et al. [13]). This method constitutes a generalization of linear decision boundaries for classification and produces nonlinear boundaries by constructing a linear boundary in a large, transformed version of the feature space. Support vector machines involve a kernel function $K(x, x')$ that computes inner products in the transformed space. Two popular choices for K in the SVM literature are the dth degree polynomial, $K(x, x') = (1+ < x, x' >)^d$ and the radial basis, $K(x, x') = \exp(-\|x - x'\|^2/c)$. Tables 1 and 2 show the misclassification error for a linear, quadratic and cubic polynomial kernel and for a radial basis kernel with default scaling parameter $c = 1$. Tables 3 and 4 show the results for different values of the scaling parameter. We can see how for the pair (Right tangent, left α) the quadratic SVM and radial basis kernel SVM with $c = 8$ performs pretty well with averaged misclassification errors equal to 0.2368 and 0.2335 respectively, for the 50% cross-validation case. For the 15% case, these errors decrease to 0.2007 and 0.2025. Figure 9 depicts the training set (74 samples), the support vectors and the non linear boundary for the quadratic and radial basis (with $c = 8$) SVM for the pair that best discrimantes, (Right tangent, Left α point). For this particular iteration, the misclassification errors were 0.2027 and 0.1486.

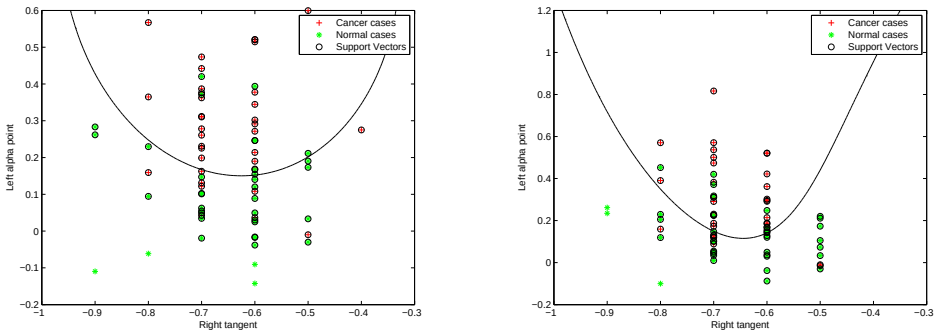

Figure 9: Training sets and nonlinear boundaries by quadratic and radial basis SVM's ($c = 8$). The circles denote the support vectors.

In addition, we also tried two regression methods for classification. Firstly, a logistic regression where the covariates or predictors (X) were the eight measures for every observation i,

Pairs of Measures	Lineal	Quad.	Decision tree	Lineal svm	Quad. svm	Poly. svm	RBF svm
(Hurst exp., left tang.)	0.2720	0.2775	0.2782	0.2741	0.2813	0.2988	0.3076
(Hurst exp., Bandwidth)	0.2704	0.2682	0.2809	0.2746	0.2638	0.2995	0.3114
(Left tang., left α)	0.2686	0.2748	0.2889	0.2660	0.2686	0.2860	0.2724
(Right tang., left α)	0.2824	0.2488	0.2844	0.2787	**0.2368**	0.2745	0.2583
(Right slope with H, left α)	0.2691	0.2714	0.3103	0.2682	0.2772	0.3290	0.3108
(Bandwidth, left α)	0.2778	0.2761	0.3229	0.2771	0.2737	0.3320	0.3525
(Bandwidth, right α)	0.2779	0.2751	0.3378	0.2766	0.2737	0.3320	0.3525

Table 1: Misclassification errors for some pair of measures and discrimination methods. Test sample = 50%.

Pairs of Measures	Lineal	Quad.	Decision tree	Lineal svm	Quad. svm	Poly. svm	RBF svm
(Hurst exp., left tang.)	0.2659	0.2695	0.2648	0.2756	0.2696	0.2787	0.2505
(Hurst exp., Bandwidth)	0.2626	0.2568	0.2877	0.2705	0.2602	0.2832	0.2802
(Left tang., left α)	0.2640	0.2760	0.2913	0.2615	0.2646	0.2497	0.3260
(Right tang., left α)	0.2869	0.2392	0.2693	0.2812	**0.2007**	0.2570	0.2357
(Right slope with H, left α)	0.2656	0.2506	0.2981	0.2572	0.2750	0.3161	0.2829
(Bandwidth, left α)	0.2670	0.2670	0.2992	0.2701	0.2588	0.3186	0.3378
(Bandwidth, right α)	0.2626	0.2657	0.3262	0.2694	0.2572	0.3193	0.3402

Table 2: Misclassification errors for some pair of measures and discrimination methods. Test sample = 15%.

and the response variable (Y_i) was 0 for the normal cases and 1 for the cancer cases, for $i = 1, \ldots, n$, where n is the sample size (148). Matrix X has got n rows and eight columns, so that X_i is the $i-th$ row, corresponding to the $i-th$ observation. The vector of parameters is denoted by β

Adding a first row of ones to matrix X, the model can be stated as,

$$\log\left(\frac{p}{1-p}\right) = X\beta \tag{4.1}$$

where the goal of the analysis were to estimate the vector p, which contains the probabilities of 1 (having cancer), given the

Pairs of Measures	$c = 2$	$c = 4$	$c = 6$	$c = 8$	$c = 10$	$c = 20$
(Hurst exp., left tang.)	0.2814	0.2549	0.2580	0.2735	0.2772	0.2779
(Hurst exp., Bandwidth)	0.2927	0.2760	0.2662	0.2664	0.2672	0.2647
(Left tang., left α)	0.3035	0.2621	0.2581	0.2714	0.2707	0.2634
(Right tang., left α)	0.2707	0.2450	0.2348	**0.2335**	0.2339	0.2381
(Right slope with H, left α)	0.3084	0.2862	0.2787	0.2763	0.2790	0.2774
(Bandwidth, left α)	0.3277	0.2894	0.2768	0.2730	0.2724	0.2725
(Bandwidth, right α)	0.3261	0.2869	0.2779	0.2747	0.2725	0.2716

Table 3: Misclassification errors for some pair of measures and different values of the scaling factor c in the radial basis for the svm. Test sample = 50%.

Pairs of Measures	$c = 2$	$c = 4$	$c = 6$	$c = 8$	$c = 10$	$c = 20$
(Hurst exp., left tang.)	0.2444	0.2387	0.2612	0.2704	0.2736	0.2717
(Hurst exp., Bandwidth)	0.2775	0.2727	0.2670	0.2631	0.2619	0.2635
(Left tang., left α)	0.2570	0.2382	0.2504	0.2495	0.2480	0.2504
(Right tang., left α)	0.2552	0.2272	0.2071	**0.2025**	0.2058	0.2130
(Right slope with H, left α)	0.3031	0.2803	0.2816	0.2730	0.2736	0.2719
(Bandwidth, left α)	0.3215	0.2662	0.2577	0.2615	0.2603	0.2614
(Bandwidth, right α)	0.3165	0.2675	0.2608	0.2576	0.2568	0.2567

Table 4: Misclassification errors for some pair of measures and different values of the scaling factor c in the radial basis for the svm. Test sample = 15%.

covariates: $p = P(Y = 1|X)$. This is given by

$$\hat{p} = \frac{e^{X\hat{\beta}}}{1 + e^{X\hat{\beta}}} \tag{4.2}$$

The second regression model we implemented was done in Bayesian fashion. We also had binary responses and the same covariates. We assumed that

$$Z_i = X_i\beta + \epsilon_i, \quad \epsilon_i \sim F, \quad i = 1, \ldots, n \tag{4.3}$$

is a multivariate regression model in which Z_i's are not observable, but the indicators $Y_i = \mathbf{1}(Z_i > 0)$ are. Then,

$$p_i = P(Y_i = 1) = P(Z_i > 0) = P(X_i\beta + \epsilon_i > 0) = 1 - F(-X_i\beta) \tag{4.4}$$

If F is normally distributed, then

$$\beta|Z,Y \sim MVN((X'X)^{-1}X'Z,(X'X)^{-1}) \tag{4.5}$$

Moreover, if Y_i and β are given, Z_i follows a truncated at 0 normal distribution with mean $X_i\beta$. Truncation is to the left if $Y_i = 1$ and to the right if $Y_i = 0$. These two conditional distributions define an easy Gibbs sampler (see Robert and Casella [28]) which gives the posterior distributions for β. We estimated β using the expected posterior values for $\beta_0, \beta_1, \ldots, \beta_8$. The probabilities p_i can be estimated then as,

$$\hat{p}_i = F(X_i\hat{\beta})$$

For every method, after having the estimations of p_i for all cases we assigned a 1 to those samples where $\hat{p}_i > 1/2$ and 0 otherwise. Then, we compared with the sample and computed the prediction errors by cross-validation. Both methods performed similarly or worse than the previous ones, with errors around 0.29.

5 Conclusions

In this work we have shown how to extend the concept of wavelet-based multifractal spectrum to the case of two dimensions. This tool, that detects different degrees of irregularity in the signal, has been widely utilized in several fields from physics to meteorology and medicine, where self-similary and fractality are involved. Many applications can be found for this 2D signal analysis tool, in particular in medicine, where radiologists and other experts have to deal with medical images.

One meaningful implication of this research applies to diagnosis of breast cancer. Current methods to detect breast tumors are costly and sometimes painful for patients. With our approach and just working with the image background we can classify patient with a reasonable misclassification error. So, our method can be seen as a "first step" in analyzing mammograms. Interesting findings were obtained in this work. For example, normal breast tissue tend to be more irregular (present a smaller Hurst exponent) than tumor affected tissue. But probably, the most surprising result was that in some way all the breast tissue in a

tumor affected patient is characterized by similar MFS measures, not only the tissue close to the tumor.

In this paper we also applied and reviewed the most common classification methods, and found that the one which best performs is the Suppor Vector machines classifier; it is the more flexible one since it allows for non-linear boundaries.

A number of extensions are possible. Firstly, and following the work in [22], we would like to study the statistical properties of the estimators defined in (3.1) and (3.2) for the case of a 2D Fractional Brownian motion, that is, asymptotic normality, bias and variance.

Another important objective is to apply our approach with different images, in some geophysical fields with satellite images or meteorological ones, and also with other medical images, related to tumors or not.

APPENDIX I

References

[1] Arnodo, A., Decoster, N. and Roux, S.G. (2000). A wavelet-based method for multifractal image analysis: I. Methodology and test applications on isotropic and anisotropic rough surfaces. *The European physical journal B*, 15, 567–600.

[2] Bocchi, L., Coppini, G., Nori, J. and Valli, G. (2004). Detection and clustered microcalcifications in mammograms using fractals models and neural networks. *Medical Engineering & Physics*, 26, 303–312.

[3] Bruce, L.M. and Adhami, R.R. (1997). Wavelet based feature extraction for mammographic lesion recognition. SPIE's International Symposium on Medical Imaging, CA, vol. 3304.

[4] Chen, CC., Daponte, JS. and Fox, MD. (1989). Fractal features analysis and classification in medical imaging. *IEEE Transactions on Medical Imaging*,8, 133–142.

[5] Curpen, B. N., Sickles, E. A., Sollitto R. A. (1995). The comparative value of mammographic screening for women 40-49 years old versus women 50-59 years old. *AJR*, 164, 1099–1103.

[6] Delbeke, L. and Abry, P. (1998). Wavelet-based estimators for the self-similar parameter of fractional Brownian motion. Submitted to *Appl. Comp. Harm. Anal.*

[7] Ellis, R. (1984). Large deviations for a general class of random vectors. *Ann. Prob*, 12,1–12.

[8] Flandrin, P. (1992). Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. Info. Theory*, 38, 910–917.

[9] Gonçalves P., Riedi, R. and Baraniuk, R. (1998). Simple statistical analysis of waveletbased multifractal spectrum estimation. In *Proceedings 32nd Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA.

[10] Hastie, T., Tibshirani, R. and Friedman, J. (2001). The elements of Statistical Learning. Data mining, Inference, and Prediction. Prentice-Hall. Usa.

[11] Heath, M., Bowyer, K.W., Kopans, D. et al,"Current status of the Digital Database for Screening Mammography", pp 457–460, *Digital Mammography*, Kluwer Academic Publishers, 1998.

[12] Heath, M., Bowyer, K., Kopans, D., Moore R., and Kegelmeyer P. Jr. (2000). The Digital Database for Screening Mammography, in *The Proceedings of the 5th International Workshop on Digital Mammography* (Toronto, Canada, June 2000), Medical Physics Publishing (Madison, WI), ISBN 1-930524-00-5.

[13] Kalisher, L. (1979). Factors influencing false negative rates in xero-mammography. *Radiology*, 133.

[14] Kestener, P., Lina, J.M., Saint-Jean, P. and Arneodo, A. (2001). Wavelet-based multifractal formalism to assist in diagnosis in digitized mammograms. *Image Anal Stereol*, 20, 169–174.

[15] Koutras, A., Christoyianni, I., Georgoulas, G. and Dermatas, E. (2006). Computer Aided Classification of Mammographic Tissue Using Independent Component Analysis and Support Vector Machines. *ICANN (2), Lecture Notes in Computer Science, Springer.* 4132, 568–577.

[16] Kuklinski, W.S. (1994). Utilization of fractal image models in medical image processing. *Fractals*, 2, 363–369.

[17] Lado, M.J., Tahoces, P.G., Mendez, A.J., Souto, M. and Vidal, J.J. (1999). A wavelet-based algorithm for detecting clustered microcalcifications in digital mammograms. *Med. Phys*, 26:7, 1294–1305.

[18] Levy, P. (1948). Processus stochastiques et mouvement Brownien. Pub-gauthier, publisher. Paris, France.

[19] Mandelbrot, B. and van Ness, J.W. (1968). Fractional Brownian motions, fractional noises and applications. *SlAM* Rev, 4:10, 422–437.

[20] Martin, J., Moskowitz, M. and Milbrath, J. (1979). Breast cancer missed by mammography. *AJR*, 132.

[21] Mavroforakis, M.E, Georgiou, H.V, Dimitropoulos, N., Cavouras, D. and Theodoridis. S. (2006). Mammographic masses characterization based on localized texture and dataset fractal analysis using linear, neural and support vector machine classifiers. *Artificial Intelligence in Medicine*, 37:2, 145–162.

[22] Moloney, K.P, Jacko, J.A, Vidakovic, B., Sainfort, F, Leonard, V.K, and Shi, B. (2006). Leveraging data complexity: Pupillary behavior of older adults with visual impairment during HCI. *ACM Trans. Comput.-Hum. Interact.*, 13:3, 376–402.

[23] Morales, C. J. (2002). Wavelet-based multifractal spectra estimation: Statistical aspects and applications. Ph.D thesis. Boston University Graduate School of Arts and Sciences.

[24] Priebe, CE., Solka, JL., Lorey, RA., Rogers, GW., Poston, WL., Kallergi, M., Quian, W., Clarke, LP. and Clark, RA. (1994). The application of fractal analysis to mammographic tissue classification. *Cancer letters*, 77, 183–189.

[25] Riedi, R. Multifractal Processes in Theory and Applications of Long-Range Dependence. (2002). Paul Doukhan, George Oppenheim, Murad S. Taqqu Editors. In-Press, 625–716.

[26] Riedi, R. (1999). Introduction to multifractals.

[27] Riedi, R., Crouse, M.S., Ribeiro, V., and Baraniuk, R.G. (1999). A multifractal wavelet model with applications to TCP network traffic. *IEEE Trans. Info. Theory* (special issue on multiscale statistical signal analysis and its applications).

[28] Robert C.P and Casella, G. (2005). Monte Carlo Statistical Methods (Springer Texts in Statistics). Springer-Verlag New York.

[29] Smart, C. R., Hendrick, R. E., Rutledge J. H., and Smith, R. A. (1995). Benefit of mammography screening in women ages 40 to 49 years: current evidence from randomized controlled trials. *Cancer*,75, 1619–1626.

[30] Vidakovic, B. (1999). Statistical Modeling by Wavelets. Wiley, NY, USA.

[31] Yoshida, H., Doi, K., Nishikawa, R.M., L.Giger, M. and A.Schmidt, R. (1996). An improved computer-assisted diagnosis scheme using wavelet transform for detecting clustered microcalcifications in digital mammograms. *Acad. Radiol.*, 3, 621–627.

[32] Zheng, B., Qian, W. and Clarke, L.P. (1996). Digital mammography: mixed feature neural network with spectral entropy decision for detection of microcalcifications. *Med. Img.*, 15, 589–597.